

文章编号:1671-251X(2017)09-0021-05

DOI:10.13272/j.issn.1671-251x.2017.09.004

智慧矿山智能决策支持技术架构设计

高文^{1,2}

(1. 中煤科工集团常州研究院有限公司, 江苏 常州 213015;

2. 天地(常州)自动化股份有限公司, 江苏 常州 213015)

摘要:在分析矿山决策支持系统应用现状的基础上,提出基于新一代大数据技术和云计算技术的智慧矿山智能决策支持技术架构。该架构采用 Hadoop 等大数据技术解决传统决策支持系统存在的无法水平扩展计算资源、响应时间长及无法把离散的管理信息数据与自动化测控数据相结合进行数据分析等问题,通过引入流式计算解决传统决策支持系统无法有效处理无界数据的问题。该架构具有可承载矿山专业业务分析模型和常规管理信息化业务分析模型的能力,可全面融合“人”、“机”、“环”、“管”相关数据,形成综合性较强的决策支持信息,为实现矿山“智慧化”提供核心技术支持。

关键词:智慧矿山;智能决策支持;技术架构;大数据;云计算;流式计算

中图分类号:TD67 文献标志码:A 网络出版时间:2017-08-28 10:24

网络出版地址:<http://kns.cnki.net/kcms/detail/32.1627.TP.20170828.1024.004.html>

Design of intelligent decision support technology architecture for wisdom mine

GAO Wen^{1,2}

(1. CCTEG Changzhou Research Institute, Changzhou 213015, China;

2. Tiandi(Changzhou)Automation Co., Ltd., Changzhou 213015, China)

Abstract: Intelligent decision support technology architecture for wisdom mine based on new generation of big data technology and cloud computing technology was put forward based on analysis of application status of mine decision support system. The architecture uses Hadoop and other big data technologies to solve problems of traditional decision support system, including unable to horizontally expand computing resources, long response time, and unable to analyze data combining with discrete management information data and automated measurement and control data; through introduction of stream calculation, solves problem that the traditional decision support systems can not effectively deal with unbounded data. The architecture has capability of carrying out professional analysis model of mine and information business analysis model for general management, it can fully integrate relevant data of "people", "machine", "ring" and "management", form a comprehensive decision support information, and provide core technical support for realization of mine intelligence.

Key words: wisdom mine; intelligent decision support; technology architecture; big data; cloud computing; stream calculation

收稿日期:2017-06-16;修回日期:2017-07-17;责任编辑:胡娴。

基金项目:天地科技股份有限公司智慧矿山专项项目(2016-ZHKSZX-03);天地(常州)自动化股份有限公司科研项目(2017GY006)。

作者简介:高文(1981-),男,山西朔州人,工程师,主要研究方向为煤矿信息化软件、监测监控软件、云计算和大数据应用等,E-mail: digital_x@163.com。

引用格式:高文.智慧矿山智能决策支持技术架构设计[J].工矿自动化,2017,43(9):21-25.

GAO Wen. Design of intelligent decision support technology architecture for wisdom mine[J]. Industry and Mine Automation, 2017, 43(9):21-25.

0 引言

当前社会已经进入了智慧化社会,而数字化矿山也正在向着智慧矿山发展。具有智能决策支持能力是智慧矿山的关键特征之一,智慧矿山系统可实现自主的智能分析,具有知识积累和模式发现能力,可以在原始数据的基础之上为决策者提供深度分析后的高价值信息,辅助矿山安全生产和经营管理。本文通过阐述智慧矿山建设现状及数据分析技术的应用现状,表明了决策支持系统还存在性能不足、数据缺失、关联性分析和综合性分析较弱的问题,针对这些问题,提出了基于新一代大数据技术的智慧矿山智能决策支持技术架构设计方案,着重介绍了架构体系的组成和关键技术实现方法。

1 智慧矿山建设现状

随着数字化矿山建设的不断深入,各科研院所、厂商和矿山企业纷纷投入到数字矿山及智慧矿山的建设和建设中,通过建设自动化监测和控制子系统、工业网络、信息化管理平台等分类项目,取得了一定的成效。虽然从事矿井数字化建设的厂家很多,但由于采矿行业的专业性相对较强,每个厂商又都只专注于自己擅长的领域,所以,只是实现了生产、安全、管理数据的集中采集、存储与展示,未考虑对数据的深度融合,在数据综合利用及如何为煤矿精细化管理提供决策支持方面考虑较少。除了部分大型矿山企业实现了传统商业智能(Business Intelligence, BI)系统的运营指标和模拟推演功能^[1],以及类似通风网络解算、灾害预警等独立专业软件,未见决策支持方面深入应用和综合平台化应用的案例。

智慧企业不是传统的数字化、信息化、智能化,它是在企业实现业务量化的基础上,将先进的信息技术、工业技术和管理技术高度融合,从而产生的一种全新的、具备自动管理能力的企业组织形态和管理模式^[2]。智慧矿山也在智慧企业的范畴中,同样应遵循上述基本规律。但在矿山行业,与一般企业不同的是,由于矿山企业生产条件复杂、环节多、不可预见因素多,难以像制造业等行业一样对需求进行实时精准的预测,建立实时的客户反馈机制^[3]来调节生产。同时,在信息技术方面,由于数字化矿山缺少完整统一的标准与规范体系支撑,使得矿山在进行类似项目建设时缺乏上层规划与总体设计,导致在实际建设过程中,系统零散、重复投资、新的信息孤岛等问题较为突出。因此,在智慧矿山建设中应把新一代大数据分析技术应用在决策支持架构

中,通过统一平台、容纳异构业务系统、增强非结构化数据处理能力等解决上述问题。

2 数据分析技术在智慧矿山中的应用

当前,我国许多矿井的信息化建设已经深入到安全生产过程的各个方面,采集到海量的传感器数据和人工录入数据后,如何充分利用这些数据并使其产生价值,从而为矿井安全生产提供一定的决策支持,是许多现代化矿井面临的问题。典型的大数据分析应用有设备健康状况分析和综采工作面分析等。设备健康状况分析:为每一种设备建立健康模型,当报警次数多、时间长时,健康值就会下降;通过对采集的数据建立一个固定时间宽度的滑动窗口,在窗口上建立相应的回归预测模型,实现对设备健康状况的预测^[4]。综采工作面分析:提供设备故障诊断、工作面顶板管理、周期来压预测、操作行为改进等分析功能。

类似这样的大数据分析应用已在部分煤矿运行,但效果尚待观察。从技术层面来看,出现的问题主要有传感器安装不足导致的监测参数缺失,日常运维没有跟上导致的传感器损坏、数据漂移、误报警等,这就使得应用在初期达到了一定的效果,后期逐步偏离目标。单纯依靠传感器获取监测数据的问题在于数据来源较为单一,数据之间没有佐证,关联分析较弱,且大量安装传感器成本较高,在非关键环节、非关键设备上常常不装传感器,导致信息缺失。因此,在智慧矿山的大数据分析中,应该将传感器数据、监控视频流、音频流和人工产生的数据相结合,共同参与到分析模型中,实现综合分析。从应用层面来看,矿山生产过程中“人”、“机”、“环”、“管”各方面在事实上是互相关联的,单独割裂开来进行分析应用必然导致结果不准确或决策参考价值较低。传统的数据分析重点放在“机”和“环”方面,对于灵活多变的“人”和“管”因素缺乏有效的结合。从大数据在其他行业的应用经验来看,人的行为数据、管理中产生的过程数据都是分析的重点。

在数据分析应用技术架构和落地实施方面,矿山专业软件带有数据分析模型和结果展示功能,这些软件基本上是自有软件架构、自有分析模型,数据互通性差,模型也无法移植到新一代通用大数据分析平台中运行;而对于管理信息系统类型的数据分析产品,大多数使用的仍然是较为传统的BI技术,通过使用数据抽取、转换、装载(Extract-Transform-Load, ETL)技术从原始数据库中抽取数据,建立数据仓库,构造多维数据集,再进行报表制作。整个过程中出现的问题:多来源的数据结构混乱,梳理清楚

需要较长的时间,沟通成本巨大;实施时间很长,从需求提出到报表制作完成时间太长,无法及时响应多变的需求;涉及到传感器产生的大批量数据时,只能定制开发相应的数据抽取和处理软件,无法建立统一的架构,也就无法有效地将自动化测控数据和管理信息系统数据相结合进行分析。

针对上述现状,在智慧矿山智能决策支持技术架构的设计中,应重点考虑数据分析过程中的数据全面性,并基于这个要求设计一种能够容纳结构化、非结构化各类数据共同参与的的分析平台。

3 智能决策支持技术架构设计

由于决策支持的业务范围较广,矿山安全、生产、经营管理等多方面都可能涉及到,所以在设计架构时需要提取公共模块做较高层次的抽象,同时架构还要容纳专业模块的多种业务数据、业务模型。从技术层面来看,智能决策支持技术架构应满足如下要求。

(1) 统一基础平台。基于 Hadoop 等新一代大数据技术,实现矿山数据抽取、存储和分析处理,实现矿山决策支持模型的运行和管理,形成云端智能决策支持承载平台。针对现有的数字化矿山系统无统一的决策支持技术架构,导致已抽取好的源数据、已形成的部分数据模型无法接入云端平台的问题,形成统一、标准化的技术架构,使矿山用户可继续利用已有的信息化资源和资产,通过升级改造接入云端,同时还可以在本地提供一部分离线智能决策分

析的功能。

(2) 统一数据描述。针对现有的数据决策分析模型无统一规范、难以在同一个平台运行、难以在云端以服务形式发布的问题,形成基于云计算架构的大数据决策分析模型承载规范,使得平台能够对模型进行统一处理、运行和发布展示;统一规范平台及其接入子系统各层级的业务范畴,提出公共的数据描述规则,按数据类型限定传输格式,理清不同抽象层次的数据范围,确定上层在抽取下层数据时的取舍规则,为源数据接入决策平台的 ETL 过程提供规范和依据。

(3) 高性能。为了实现海量数据条件下的实时分析要求,需要内存计算、列计算、分布式计算等技术来支持,达到对于每个分析主题均能在最短时间内响应用户查询的效果。采用 Spark 等新一代数据处理框架可以有效地解决传统数据分析技术性能不足、无法水平扩展计算资源等问题。

(4) 自助服务能力。在数据分析时,用户可以自由地对数据进行 ETL 处理,可以随意切换维度,进行无限层次的透视分析,均为可视化操作,无须编辑代码和脚本;对于常规的结构化数据,从数据抽取、数据清洗到多维数据库表,再到前端的多维分析都提供一体化的自助服务,加速从需求提出到结果展示的过程,快速迭代。

依据上述要求,智慧矿山智能决策支持技术架构如图 1 所示。

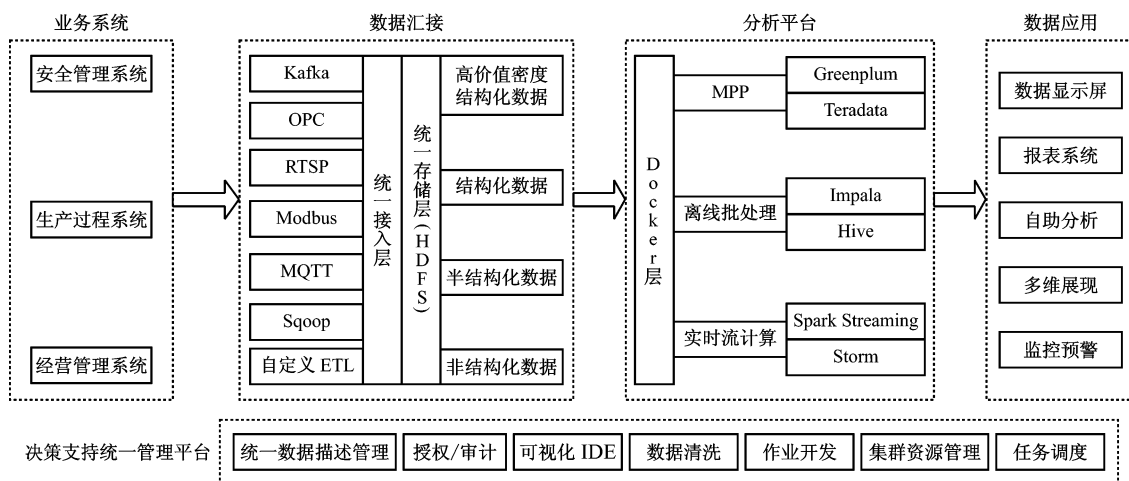


图 1 智慧矿山智能决策支持技术架构

Fig. 1 Intelligent decision support technology architecture for wisdom mine

整体架构分为业务系统、数据汇接、分析平台、数据应用和决策支持统一管理平台 5 个部分。

(1) 业务系统。业务系统包括矿山企业已建成的安全管理类、生产过程类和经营管理类系统,既有一般的管理信息化软件系统,也有工业自动化控制

系统、监测系统和视频监控系统等,这些系统是决策支持技术架构中的原始数据来源。

(2) 数据汇接。数据接入方面在传统 BI 和数字矿山等系统的建设过程中已经发展得较为成熟,但存在视频流需要单独处理、新一代物联网协议支

持度不佳等问题。智能决策支持技术架构通过增加统一接入层,把各种类型的协议、接入方式统一在一个平台上处理;使用 Kafka 中间件解决高吞吐量条件下可靠的消息订阅/发布问题,采用消息队列遥测传输(Message Queuing Telemetry Transport, MQTT)解决物联网设备接入问题;OPC 和 Modbus 协议用于连接传统的工业自动化系统或设备,使用开源工具 Sqoop 或自定义 ETL 组件抽取传统关系型数据库和文本类型的数据。数据汇接过来后,保存在统一的 Hadoop 分布式文件系统(Hadoop Distributed File System, HDFS)中。基于性能方面的考虑,可按高价值密度结构化数据、结构化数据、半结构化数据和非结构化数据来分区保存。

(3) 分析平台。为了承载多种数据分析组件和方法,通过增加基于 Docker 技术的容器层,可以在统一的基础计算平台上同时运行大量异构分析业务应用。对于高密度价值的结构化数据,可使用大规模并行处理(Massive Parallel Process, MPP)类型的数据分析工具来进行处理,如 Greenplum 或 Teradata,这样可以有效解决自助分析过程中的响应时间问题;对于海量的结构化和半结构化数据,采用基于 Hadoop 框架的上层组件 Impala 或 Hive 进行分析;对于无界的非结构化数据或测控数据,采用实时流计算工具 Spark Streaming 或 Storm 来处理,可以提供不间断的事件触发机制和滑动窗口数据分析功能。上述不同的组件和工具可以提供完整地处理异构数据、快速构建面向主题的数据仓库、高效分析数据间关联关系和准确描述数据相关性^[5]等功能,并且支持去中心化协议,选出主节点以后,再结合中心化副本控制协议^[6]完成系统整体的分布式节点管理。

(4) 数据应用。架构支持常见的数据显示屏展示、报表系统、多维展现和监控预警等数据应用,同时支持在线的自助分析功能,可以快速提取数据、快速构建查询和生成图表,过程中无需软件开发人员编写代码,通过可视化方式完成操作。

(5) 决策支持统一管理平台。智能决策支持架构的整体管理由统一的管理平台完成,其中统一数据描述管理用于解决异构数据源对业务数据描述不一致的问题,授权和审计提供了细粒度的权限管理和事务日志存档功能,可视化集成开发环境(Integrated Development Environment, IDE)用于支持业务模型开发和自助分析,数据清洗提供了矿山行业信息化系统常用的噪声数据过滤功能,作业开发用于编排数据分析事务过程,集群资源管理实现内存、CPU、网络资源和磁盘 I/O 等计算资源的

分配和回收功能,任务调度提供业务分析应用的排队、优先级等调度管理功能。

4 关键技术

4.1 异构业务承载技术

为了兼容不同的业务分析需求,架构提供了一种支持多业务模型的承载容器,容器层对上为应用层提供统一的数据接口,对内封装不同业务的数据和操作,对下提供基础计算资源的统一管理。容器技术指的是把业务处理功能打包在一个类似“集装箱”的环境中,与系统内其他进程相互隔离,互不影响。与常规虚拟机的机制有所区别的是,当容器启动时,仅是通过进程间调度,而不需要引导整个系统。通过使用轻量化的 Docker 容器承载矿山不同的业务计算模型,构建一个将不同的数据分析模型、工具或系统整合在一起的大数据分析处理平台,允许不同的大数据分析应用在一个集群内共享计算资源,但在逻辑上又互相隔离,保证了异构业务正常运行^[7-9]。

4.2 流式计算技术

针对矿山数据的多样化特性,特别是矿山测控系统产生的实时数据的存储和管理问题,可引入流式计算方案,以应对高速时序数据的在线模型分析和运算。流式计算是一类针对流数据的实时计算方法,可有效地缩短数据计算延时,实时化计算逻辑,按照时间顺序对数据提供订阅和消费功能,数据流长久且持续地进入流计算系统,一旦有新的流数据进入,流式计算立刻发起并进行一次计算任务。流式计算可执行响应函数、聚合和滚动指标等方面的计算,满足了智慧矿山中测控数据、视频数据和音频数据的处理和分析要求。

4.3 数据流降载技术

在处理生产过程自动化数据、视频监控(可变码率)数据时,常常出现瞬时突发大流量数据的情况,通过缓存+数据流降载技术可保障系统稳定运行。当数据流大于系统处理能力时,首先填充缓存等待后续处理,如果出现缓存不足的情况,则需要采用降载技术。降载技术是在尽可能保留数据特征的同时,删除冗余数据,这种技术会尽量减小对数据流最终处理的影响,具体方法包括直方图降载方法^[10]和模式特征保持降载方法^[11]等。在降载过程中,如果平台中新的计算资源分配完成,就可以关闭降载。降载技术可作为提高系统可用性的一种应急机制。

4.4 数据隔离和计费技术

新一代大数据的基础组件侧重于完成数据的存储和运算,数据的隔离不是其关注的重点。而上层

应用更关注数据的使用和业务逻辑的实现,对于缺失的数据隔离功能,可在存储层和容器层共同配合解决。数据隔离可使用多种方法来实现,在数据存储层面,可以按用户角色给数据打上标签,粒度可细化到行、列级别;在容器层面,可以为用户分配互相隔离的轻量级计算容器,同一用户仅能在容器内部执行业务计算,而又可以共享基础计算资源。

作为通用的业务容器层,在访问计费方面提供的是基础性数据,例如用户停留时间、用户基本信息、本次访问的数据量、计算所用集群节点 CPU 核数、内存使用峰值等。这些基础数据通过微服务^[12-13]的形式对外发布,可由计费组件获取并按预定的规则计算费用。

5 结语

智慧矿山智能决策支持技术架构解决了传统决策支持系统业务分散、模型不能共享、数据处理能力不足、实施成本高、门槛高等问题,基于云服务模式的智能决策应用承载平台使得矿山具备了对不同业务的决策支持系统进行顶层规划、统一实施的能力。后续在智慧矿山项目中还需要对架构中的视频处理过程做进一步优化,例如从视频流中提取有效的、正确率较高的结构化数据的功能尚待完善;Hadoop 体系也在不断演化发展,大量敏捷 BI 产品不断涌现,在决策支持的人机交互方面可引入虚拟现实(Virtual Reality, VR)、增强现实(Augmented Reality, AR)和混合现实(Mixed Reality, MR)技术,进一步降低多维度信息的理解门槛。

参考文献(References):

- [1] 韩建国. 神华智能矿山建设关键技术研发与示范[J]. 煤炭学报, 2016, 41(12): 3181-3189.
HAN Jianguo. Key technology research and demonstration of intelligent mines in Shenhua Group[J]. Journal of China Coal Society, 2016, 41(12): 3181-3189.
- [2] 涂扬举, 郑小华. 建设智慧企业, 实现自动管理[J]. 清华管理评论, 2016(10): 29-37.
- [3] 李梅, 杨帅伟, 孙振明, 等. 智慧矿山框架与发展前景研究[J]. 煤炭科学技术, 2017, 45(1): 121-128.
LI Mei, YANG Shuaiwei, SUN Zhenming, et al. Study on framework and development prospects of intelligent mine[J]. Coal Science and Technology, 2017, 45(1): 121-128.
- [4] 王智峰, 屈凡非, 田建军, 等. 基于海量数据分析的煤矿生产辅助决策支持系统的设计[J]. 工矿自动化, 2011, 37(10): 22-25.
WANG Zhifeng, QU Fanfei, TIAN Jianjun, et al. Design of auxiliary decision support system of coal mine production based on mass data analysis[J]. Industry and Mine Automation, 2011, 37(10): 22-25.
- [5] 吕佑龙, 张洁. 基于大数据的智慧工厂技术框架[J]. 计算机集成制造系统, 2016, 22(11): 2691-2697.
LYU Youlong, ZHANG Jie. Big-data-based technical framework of smart factory[J]. Computer Integrated Manufacturing Systems, 2016, 22(11): 2691-2697.
- [6] 朱进云. 大数据架构师指南[M]. 北京: 清华大学出版社, 2016.
- [7] 李丽新. 基于 Mesos 和 Docker 构建大数据平台技术研究与应用[J]. 吉林建筑大学学报, 2016, 33(5): 81-84.
LI Lixin. Research and application of big data platform based on mesos and docker[J]. Journal of Jilin Jianzhu University, 2016, 33(5): 81-84.
- [8] 赵乐乐, 黄刚, 马越. 基于 Docker 的 Hadoop 平台架构研究[J]. 计算机技术与发展, 2016, 26(9): 99-103.
ZHAO Lele, HUANG Gang, MA Yue. Research on Hadoop platform based on Docker[J]. Computer Technology and Development, 2016, 26(9): 99-103.
- [9] 陈存香, 陈清金, 张岩. Hadoop 与 Docker 技术的融合[J]. 邮电设计技术, 2015(5): 5-8.
CHEN Cunxiang, CHEN Qingjin, ZHANG Yan. Integration technology between Hadoop and Docker[J]. Designing Techniques of Posts and Telecommunications, 2015(5): 5-8.
- [10] 魏霞, 李国徽. 基于直方图的数据流降载策略研究[J]. 华中科技大学学报(自然科学版), 2014, 42(9): 24-27.
WEI Xia, LI Guohui. Histogram strategy for load shedding over data streams[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2014, 42(9): 24-27.
- [11] 苗高杉, 李红燕. 基于模式特征保持的数据流降载方法[J]. 计算机研究与发展, 2010, 47(增刊 1): 309-315.
MIAO Gaoshan, LI Hongyan. Pattern feature preserving strategy for load shedding over data streams[J]. Journal of Computer Research and Development, 2010, 47(S1): 309-315.
- [12] 杜圣东, 杨燕, 滕飞. 交通大数据: 一种基于微服务的敏捷处理架构设计[J]. 大数据, 2017(3): 53-67.
DU Shengdong, YANG Yan, TENG Fei. Traffic big data: an agile architecture design based on micro service[J]. Big Data Research, 2017(3): 53-67.
- [13] 张晶, 王琰洁, 黄小锋. 一种微服务框架的实现[J]. 计算机系统应用, 2017, 26(4): 82-86.
ZHANG Jing, WANG Yanjie, HUANG Xiaofeng. Implementation of microservice architecture[J]. Computer Systems & Applications, 2017, 26(4): 82-86.